**Children’s Active & Independent Travel**

*Data Assembly Notes*

Spring 2024

\_\_\_

*Contents:*

1. Catalogue of variables in the National Household Travel Survey from years 2009, 20017, and 2022 (used for this analysis)
2. Variables and basis statistics for each of the two assembled datasets:
   1. School Trips (2009, 2007, 2022)
   2. All trips (2009, 2007, 2022)
3. Data notes and next steps

***1. National Household Travel Survey Variables used for this analysis and differences between years***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **2017 data**  ***From NHTS*** | | **2009 data**  ***From NHTS*** | **2022 data**  ***From NHTS*** |
|  | **Sample** | | | | |
|  | # of trips | 923,572 |  | 1,167,321 | 31,074 |
|  | # of individuals | 264,234 |  | 308,901 | 16,997 |
|  | # of households | 129,696 |  | 150,147 | 7,893 |
|  | **Variables** | | | | |
| Trip characteristics | Trip distance in miles | **TRPMILES**  n, 0 – 9621 | Only include trips < 0 miles; converted to KM | Same as 2017 | Same as 2017 |
| Trip origin purpose | **WHYFROM**  c | 7 = Transfer, 8 = Attend school as a student | Different coding:  21 = School as a student; no equivalent for “Change type of Transport” | Different coding:  6 = Attend school as a student, 9 = Transfer |
| Trip destination purpose | **WHYTO**  c | 7 = Transfer, 8 = Attend school as a student | Different coding:  21 = School as a student; no equivalent for “Change type of Transport” (new for 2017) | Different coding:  6 = Attend school as a student, 9 = Transfer |
| Trip number in the day | **TDTRPNUM**  c, 1- 50 |  | Same as 2017 | Named differently: TRIPID |
| Number of people on trip, incl. respondent | **NUMONTRP**  n, 1 - 401 |  | Same as 2017 | Same as 2017 |
| Number of non-HH people on trip, derived | **NONHHCNT**  n, 1 - 400 |  | Same as 2017 | Same as 2017 |
| HH person identifier on trip | **ONTD\_P1-13**  C, yes = 1, no = 2 |  | Same as 2017 | Same as 2017 |
| Trip mode, derived | **TRPTRANS**  c | 3-6 = car (car, SUV, Van, Pickup truck); 10 = school bus; 1 = walk; 2 = bike; 11 – 16 = transit (bus, paratransit, dial-a-ride, shuttle, rail, subway);  8 = motorcycle  97 = unspecified  Suggest including ’18 Rental Car’ in ‘Car’ category  Suggest including 20 Ferry in ‘Transit’ category | Different categorization  01-04 = car (car, Van, SUV, pick-up truck); 07 = motorcycle; 09,10,12-18 = transit (bus, train, trolley), 22 = bike, 23 = walk, 24 = special transit, 97 = other; 11 = school bus  If including 18 Rental Car, then add 05 ‘Other Truck’  If including 20 Ferry, then add 20 Ferry | Different categorization  01-04 = car (car, Van, SUV, pick-up truck); 07 = motorcycle; 08, 10-13,17,22 = transit (bus, train, trolley), 18,19 = bike/e-scooter, 20 = walk, 21 = other; 09 = School bus |
| Trip end time | **ENDTIME**  c, HHMM | Only include trips < 1000 (i.e., ends before 10am) | Same as 2017 | Same as 2017 |
| Location characteristics | Pop density – origin (census block group) | **OBPPOPDN**  c, 50, 300, 750, 1500, 3000, 7000, 17000, 30000; |  | No equivalent in 2009 data; only have population density for household’s home location (same numerical categories) - HBPPOPDN | No equivalent in 2022 data; only have URBAN variables characterizing nature (urban/rural) of region where home address is |
| Pop density – destination (census block group) | **DBPPOPDN**  c, 50, 300, 750, 1500, 3000, 7000, 17000, 30000 |  | No equivalent in 2009 data; only have population density for household’s home location (same numerical categories) - HBPPOPDN | No equivalent in 2022 data; only have URBAN variables characterizing nature (urban/rural) of region where home address is |
| Distance from home to school | **DISTTOSC17**  n, 0-4,500 |  | Categorical variable in 2009 data; ranges e.g., less than 1 mile, more than 2 miles | No equivalent in 2022 data (only distance from home to work) |
| Traveler characteristics | Person ID | **PERSONID**  c, 1-11 |  | Same as 2017 (1-13) | Same as 2017 (1-10) |
| Traveler age | **R\_AGE**  n, 5-92 |  | Same as 2017; however 89+ captured as 92 | Same as 2017 |
| Traveler gender | **R\_SEX\_IMP**  c, Male or Female only |  | Captured in R\_SEX (only male/female) | Same as 2017 |
| Worker status | **WORKER**  c, Yes/No, not ascertained, skip |  | Same as 2017 | Same as 2017 |
| Driver status, derived | **DRIVER**  c, Yes/No, skip |  | Same as 2017 | Same as 2017 |
| Household characteristics | House ID | **HOUSEID** |  | Same as 2017 | Same as 2017 |
| Household income | **HHFAMINC**  c, 11 income categories | Mid-point of income category; if greater than $200k = $250k (code lines 382 – 393) | Different to 2017: 18 income categories | Same as 2017 |
| Household size | **HHSIZE**  n, 1-13 |  | Same as 2017 (1-14) | Same as 2017 (1-10) |
| Number of household vehicles | **HHVEHCNT**  n, 1-12 |  | Same as 2017 (1-27) | Same as 2017 (0-17) |
| Number of household drivers | **DRVRCNT**  n, 0-9 |  | Same as 2017 (0-9) | Same as 2017 (0-7) |

***2. Descriptions of the two assembled datasets***

*School trips – 2009, 2017, 2022*

**Predictor variables:**

* `log\_inc\_k`: Income (log-transformed, continuous)
* `veh\_per\_driver`: Vehicles per driver
* `non\_work\_mom`: Presence of non-working adult woman in household
* `non\_work\_dad`: Presence of non-working adult man in household
* `age`: Age
* `female`: Sex
* `has\_lil\_sib`: Presence of younger child in household
* `has\_big\_sib`: Presence of older child in household
* `log\_distance`: Distance
* `log\_density`: Population density of home block group
* `density\_alt`: Categorical density variable (urban or rural)

**And the outcome variable, `mode\_ind`, which can take any of the following values:**

* 18: Active mode accompanied by a household adult
* 28: Active mode unaccompanied by a household adult
* 17: In a car accompanied by a household adult
* 27: In a car unaccompanied by a household adult

We assume that all of the above alternatives are available for all respondents, so we will also create the following availability variables, and set the values to 1/TRUE for all respondents.

* av\_kid\_car
* av\_par\_car
* av\_kid\_act
* av\_par\_act

|  |  |  |  |
| --- | --- | --- | --- |
|  | **2009** | **2017** | **2022** |
| **# of trips in original data set** | 1,167,321 | 923, 752 | 7,893 |
| *# of trips excluded (school bus)* | 15,146 | 11,313 | 650 |
| *# of trips excluded (transit)* | 10,540 | 13,415 | 354 |
| **# of trips in sample** | **2834** | **2076** | **103** |

*All trips – 2009, 2017, 2022*

**Predictor variables:**

* `purpose`: indicates the trips destination or purpose
* `school`: whether this was a school trip or not
* `log\_inc\_k`: Income (log-transformed, continuous)
* `veh\_per\_driver`: Vehicles per driver
* `non\_work\_mom`: Presence of non-working adult woman in household
* `non\_work\_dad`: Presence of non-working adult man in household
* `age`: Age
* `female`: Sex
* `has\_lil\_sib`: Presence of younger child in household
* `has\_big\_sib`: Presence of older child in household
* `log\_distance`: Distance
* `log\_density`: Population density of home block group
* `density\_alt`: Categorical density variable (urban or rural)

**And the outcome variable, `mode\_ind`, which can take any of the following values:**

* 18: Active mode accompanied by a household adult
* 28: Active mode unaccompanied by a household adult
* 17: In a car accompanied by a household adult
* 27: In a car unaccompanied by a household adult

We assume that all of the above alternatives are available for all respondents, so we will also create the following availability variables, and set the values to 1/TRUE for all respondents.

* av\_kid\_car
* av\_par\_car
* av\_kid\_act
* av\_par\_act

|  |  |  |  |
| --- | --- | --- | --- |
|  | **2009** | **2017** | **2022** |
| **# of trips in original data set** | 1,167,321 | 923, 752 | 7,893 |
| *# of trips that were transfers* | 0 | 17,339 | 328 |
| **# of trips in sample** | **24,196** | **13,508** | **481** |

***3. Data notes and next steps***

**School Trips:**

* **Line 182:** I included ‘rideshare’ (18) in the ‘car’ category. This was only an option in 2017
* **Line 186/223/259:** I added the URBRUR categorical variable for each year to indicate density (urban or rural) given 2022 does not have a continuous density variable.
  + **This variable is titled density\_alt and added to the final ‘trips’ table**
* **Line 237:** I also approximated a continuous density variable for 2022 > if URBRUR = urban, density = 7000 and if URBRUR = rural, density = 50. The values are based on the numeric categories in the 2009 and 2017 density variables
* **Line 380-290:** To convert 2022 income to 2017, I multiplied by 0.86 (based on <https://data.bls.gov/cgi-bin/cpicalc.pl>, January)
* **[ACTION] Line 390:** The highest income bracket for 2022 still needs to be determined (used the same as 2017 for now)

**All trips:**

* **As above,** I’ve used two types of density variables
* **Line 600>:** To convert 2022 income to 2017, I multiplied by 0.86 (based on <https://data.bls.gov/cgi-bin/cpicalc.pl>, January)
* **[ACTION] Line 640:** The highest income bracket for 2022 still needs to be determined (used the same as 2017 for now)
* **[ACTION] Some predictor variables may still need to be added in the all-**trips data; for now, the only additional predictor (different to school-trips) is “school-trip” indicating whether the trip is a school trip or not. Previously included variables:
  + school\_dist
  + avg\_trip\_dist
  + n\_non\_school\_trips
  + had\_school: binary (true, false)